Submitted to Physical Review E

# Chaos for Liouville probability densities

Rüdiger Schack* and Carlton M. Caves
*Center for Advanced Studies, Department of Physics and Astronomy,*
*University of New Mexico, Albuquerque, NM 87131-1156*
(February 5, 2008)



## Abstract

Using the method of symbolic dynamics, we show that a large class of classical chaotic maps exhibit *exponential hypersensitivity to perturbation*, i.e., a rapid increase with time of the information needed to describe the perturbed time evolution of the Liouville density, the information attaining values that are exponentially larger than the entropy increase that results from averaging over the perturbation. The exponential rate of growth of the ratio of information to entropy is given by the Kolmogorov-Sinai entropy of the map. These findings generalize and extend results obtained for the baker's map [R. Schack and C. M. Caves, Phys. Rev. Lett. **69**, 3413 (1992)].

## I. INTRODUCTION

Chaos in Hamiltonian systems is usually defined in terms of trajectories of phase-space points. The Lyapunov exponent describes how initially close trajectories diverge exponentially [1]. The Kolmogorov-Sinai (KS) entropy measures the rate at which information about the initial phase-space point must be supplied in order to predict the coarse-grained behavior of a trajectory at a later time [2,3].

Signatures of chaos are less obvious if attention is shifted from the time evolution of phase-space points to the time evolution of probability densities, governed by the Liouville equation. If the distance between two densities is defined in terms of an overlap integral, there is no exponential divergence of initially close densities since the overlap integral is constant in time ("Koopman's theorem" [4,5]). Furthermore, as a direct consequence of Koopman's theorem, if one is given the Hamiltonian and the initial density to a certain accuracy, then no additional information is needed to predict the density at all later times $t$ to the same accuracy, except for a negligible amount of information needed to specify the time $t$ [6,7]. This means that the popular information-theoretic interpretation [3] of chaos via the KS entropy does not apply to Liouville densities.

In this paper we show that there is an information-theoretic way to characterize chaos for Liouville densities in systems with a positive KS entropy. In particular, we show that a large class of Hamiltonian systems with positive KS entropy display an exponential hypersensitivity to perturbation. We have investigated hypersensitivity to perturbation previously [8–12],

both for classical and quantum systems, and have characterized it is as a rapid increase with time of the information needed to describe the perturbed time evolution of the system state (Liouville density for classical systems, state vector for quantum systems), the information attaining values much larger than the entropy increase that results from averaging over the perturbation.

Here we formulate the concept of hypersensitivity to perturbation more precisely. We consider the amount of information needed to keep track of the perturbed time evolution to a level of accuracy that keeps the increase of system entropy below a certain "tolerable" level. This information should be compared to the entropy reduction it buys, i.e., to the difference between the entropy increase that results from averaging over the perturbation and the tolerable entropy increase. We characterize hypersensitivity to perturbation in terms of the ratio of information to entropy reduction. A system displays *hypersensitivity to perturbation* if the ratio grows rapidly with time, becoming much larger than unity, for almost all values of the tolerable entropy; a system displays *exponential hypersensitivity to perturbation* if the ratio grows exponentially. We show that a large class of Hamiltonian systems with positive KS entropy display exponential hypersensitivity to perturbation, with the exponential growth rate given by the KS entropy. This result establishes a direct connection between measures of chaos based on trajectories and our information-theoretic characterization for Liouville densities.

There are at least two important motivations for investigating signatures of chaos in Liouville densities. One motivation comes from the tricky question of how to characterize quantum chaos. In quantum mechanics, trajectories of state vectors show no sensitivity to initial conditions because the Schrödinger equation is linear and preserves the inner product. This argument does not prove, however, that there is no chaos in quantum mechanics [13], because the Liouville equation, like the Schrödinger equation, is linear and preserves the overlap between densities, yet any chaotic classical Hamiltonian system can be described by a Liouville equation. Furthermore, the classical analog of a quantum state vector is not a point in classical phase space, but a Liouville density [5,9]. In contrast to the above-mentioned characterizations of classical chaos in terms of phase-space trajectories, a characterization of classical chaos in terms of Liouville densities can be expected to have a straightforward generalization to quantum systems [8]. We have indeed found that hypersensitivity to perturbation is present in quantum systems [11,12].

The other main motivation for studying chaos in Liouville densities lies in the central role Liouville densities play in statistical mechanics. The connection of the present work with statistical mechanics is outlined in Sec. II. In Sec. III we give a precise definition of hypersensitivity to perturbation. Section IV reviews the method of symbolic dynamics. In Sec. V, the heart of the paper, we apply the method of symbolic dynamics to prove that a large class of perturbed chaotic systems display exponential hypersensitivity to perturbation. In Sec. VI we distill the essence of the symbolic-dynamics analysis to develop a simple, heuristic picture of hypersensitivity to perturbation, which explains why chaotic systems exhibit exponential hypersensitivity to perturbation and regular, or integrable systems do not. A reader not interested in the details of the symbolic dynamics might profitably skip Secs. IV and V and proceed directly to Sec. VI.

## II. CONNECTION WITH STATISTICAL MECHANICS

In statistical mechanics the exact point a system occupies in phase space typically is not known. The predictions of classical statistical mechanics are derived from a Liouville probability density $\rho(x)$ on phase space, which describes incomplete knowledge of the system's phase-space point $x$ and which is the mathematical representation of a system state. The entropy (in bits) of a system state $\rho(x)$, also called the Gibbs entropy or fine-grained entropy, is defined as

$$H = -\int d\Gamma(x)\, \rho(x) \log_2[\rho(x)] \;, \tag{2.1}$$

where $\Gamma(x)$ is the standard phase-space measure. (The use of base-2 logarithms here and throughout this paper means that entropy and information are measured in bits.) Since the Gibbs entropy is formally identical to Shannon's [14] statistical measure of information, entropy can be interpreted as the amount of information missing toward a complete specification of the system. The classical entropy is defined up to an arbitrary additive constant, reflecting the fact that an infinite amount of information is needed to give the exact location of a point in phase space.

As a consequence of Liouville's theorem, the entropy remains constant under Hamiltonian time evolution. We adopt here the Bayesian, or information-theoretic, approach to statistical mechanics [15–17], according to which the constancy of the Gibbs entropy is an expression of the fact that no information about the initial Liouville density is lost under Hamiltonian time evolution.

The Bayesian approach to statistical mechanics is connected with thermodynamics in the following way: Assume there is a heat reservoir at temperature $T$, with which all energy in the form of heat must ultimately be exchanged, possibly by using intermediate steps such as storage at some other temperature; then each bit of missing information about the system state reduces by the amount $k_B T \ln 2$ the energy that can be extracted from the system in the form of useful work. The Bayesian approach can thus be summarized in two statements: (i) entropy is missing information—a mathematical statement; (ii) each bit of missing information costs $k_B T \ln 2$ of useful work—this is the physics.

Since entropy is a measure of missing information, entropy increases if information about the system is lost. There are two main mechanisms leading to information loss (as noted above, Hamiltonian time evolution is *not* such a mechanism): deliberate discarding of information and loss of information through interaction with an incompletely known environment.

Deliberate discarding of information was used by Jaynes [15–17] to derive traditional thermodynamics. Jaynes showed how equilibrium thermodynamics follows effortlessly from the Liouville equation if the only information retained is the values of the macroscopic variables defining a thermodynamic state. In Jaynes's approach, irrelevant information is discarded by means of the principle of maximum entropy. Another example is the derivation of the Boltzmann equation [18]; here information about correlations between particles is discarded as irrelevant.

In contrast to these examples where information is discarded deliberately, an actual loss of information can occur in a system that, rather than being perfectly isolated, interacts with an incompletely known environment. The interaction with the environment leads to a perturbed time evolution of the system. Predictions for the system alone are made by tracing

out the environment—i.e., by averaging over the perturbations—which generally leads to an entropy increase. This approach was pioneered by Borel [19,20]. The entropy increase of the system due to the interaction with the environment is a result of the environment's being in an at least partially unknown state. If suitable information about the environment is obtained, the increase in system entropy can be reduced or, if sufficient information is obtained, prevented entirely. Averaging over the perturbing environment is usually justified by arguing that it is impossible in practice to control the environment.

In this paper we go beyond the pragmatic argument that controlling the interaction with the environment is impossible in practice. We show how the information-theoretic approach to statistical mechanics leads naturally to a quantitative measure of how hard it is to keep the entropy of the system from increasing by gathering information about the environment. The key to quantifying the difficulty of controlling the interaction with the environment is Landauer's principle [21,22], which assigns a thermodynamic cost to information. According to Landauer's principle, in the presence of a heat reservoir at temperature $T$, not only does each bit of *missing* information have a free-energy cost of $k_B T \ln 2$, but each bit of information that is *acquired* has the same free-energy cost of $k_B T \ln 2$. This cost, called the *Landauer erasure cost*, is paid when the acquired information is erased. Acquired information can be quantified by algorithmic information [6,7,23–25]; roughly speaking, the algorithmic information in an observational record is the length in bits of the shortest record having the same information content.

The question of how hard it is to reduce the system entropy by controlling the environment can now be given a quantitative form: "How big is the Landauer erasure cost of the information *about the environment* which is needed to reduce the increase of system entropy by a certain amount?" In the next section we give a mathematical formulation of this question. The later parts of this paper are devoted to showing that the answer can be used to characterize chaos.

## III. HYPERSENSITIVITY TO PERTURBATION

Consider a classical Hamiltonian system initially described by a Liouville density $\rho(x, t = 0)$ on phase space. The initial entropy is

$$H_0 = -\int d\Gamma(x)\, \rho(x, t = 0) \log_2[\rho(x, t = 0)] \,, \tag{3.1}$$

where $\Gamma(x)$ is the standard phase-space measure. By solving the Liouville equation, one obtains the density $\rho(x, t)$ at time $t$. According to Liouville's theorem, the entropy remains unchanged—the information about the initial density is preserved under Hamiltonian time evolution.

Now assume that the system is coupled to an incompletely known environment in such a way that the interaction can be described as an energy-conserving, typically time-dependent perturbation of the system Hamiltonian. The system's interaction with the environment is thus described by a stochastic Hamiltonian. We denote the perturbed system state at time $t$ by $\rho_y(x, t)$ where $y$ labels the particular realization of the stochastic perturbation or *perturbation history*. The possible perturbation histories $y$ are distributed according to a probability measure $\gamma(y)$. This description in terms of a stochastic system Hamiltonian

applies when the system is coupled to conserved quantities of the environment. The values of the conserved environment quantities label the perturbation histories $y$, and the probability measure $\gamma(y)$ is the probability measure for the conserved environment quantities.

For each perturbation history $y$, the entropy of the density $\rho_y(x,t)$ is equal to the initial entropy $H_0$. Averaging over all possible perturbation histories leads to an average density

$$\bar{\rho}(x,t) = \int d\gamma(y)\, \rho_y(x,t) \ , \tag{3.2}$$

with entropy

$$\bar{H} = -\int d\Gamma(x)\, \bar{\rho}(x,t) \log_2[\bar{\rho}(x,t)] \equiv H_0 + \Delta H_{\mathcal{S}} \ , \tag{3.3}$$

where $\Delta H_{\mathcal{S}} \geq 0$ is the entropy increase due to averaging over the incompletely known environment. That $\Delta H_{\mathcal{S}} \geq 0$ follows from the concavity of the entropy: the entropy of an average distribution is greater than or equal to the average entropy of the distributions that contribute to the average.

Now assume, in accordance with the discussion of Sec. II about gathering information from the environment, that an arbitrary measurement, with discrete possible outcomes labeled by integers $b$, is performed on the environment. The outcome $b$ has conditional probability $p_{b|y}$, given the perturbation history $y$, and hence has unconditioned probability

$$p_b = \int d\gamma(y)\, p_{b|y} \ . \tag{3.4}$$

The Liouville density for the system state conditional on outcome $b$ we denote by

$$\rho_b(x,t) = \frac{1}{p_b} \int d\gamma(y)\, \rho_y(x,t) p_{b|y} = \int d\gamma(y|b)\, \rho_y(x,t) \ , \tag{3.5}$$

where $\gamma(y|b)$ is the probability measure for the perturbation histories conditional on outcome $b$. It follows immediately that

$$\sum_b p_b \rho_b(x,t) = \int d\gamma(y)\, \rho_y(x,t) = \bar{\rho}(x,t) \ . \tag{3.6}$$

We denote by

$$\Delta H_b = -\int d\Gamma(x)\, \rho_b(x,t) \log_2[\rho_b(x,t)] - H_0 \geq 0 \tag{3.7}$$

the change in system entropy conditional on the measurement outcome $b$, where the inequality follows from applying concavity to Eq. (3.5), and by

$$\overline{\Delta H} = \sum_b p_b \Delta H_b \leq \Delta H_{\mathcal{S}} \tag{3.8}$$

the average conditional entropy change, where the inequality follows from applying concavity to Eq. (3.6). Finally, we denote by

$$\overline{\Delta I} = -\sum_b p_b \log_2 p_b \tag{3.9}$$

the average information needed to specify the measurement outcome $b$. Actually, Eq. (3.9) is only a lower bound to the average algorithmic information needed to specify the measurement outcome $b$, but it can be shown to be an extremely tight lower bound [25]. An immediate consequence of the definition of entropy is that

$$\overline{\Delta I} + \overline{\Delta H} \geq \Delta H_{\mathcal{S}} \ , \tag{3.10}$$

with equality holding if and only if the densities $\rho_b(x,t)$ are disjoint.

Suppose now that one wants to limit the entropy increase of the system to a certain *tolerable amount* $\Delta H_{\rm tol}$. Then the minimum amount of information about the perturbing environment needed to keep the system entropy from increasing by more than $\Delta H_{\rm tol}$ can be written as

$$\Delta I_{\rm min} \equiv \inf_{\overline{\Delta H} \leq \Delta H_{\rm tol}} \overline{\Delta I} \ , \tag{3.11}$$

where the infimum is taken over all possible measurement schemes for which the average conditional entropy increase does not exceed $\Delta H_{\rm tol}$. In other words, $\Delta I_{\rm min}$ is the information about the environment that it takes to lower the entropy increase of the system from $\Delta H_{\mathcal{S}}$ (the increase due to averaging over the perturbation) down to $\Delta H_{\rm tol}$; i.e., $\Delta I_{\rm min}$ is the minimum information about the environment needed to reduce the system entropy by an amount $\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}$. As a consequence of Eq. (3.10), it is a general theorem—and an expression of the second law—that

$$\Delta I_{\rm min} \geq \Delta H_{\mathcal{S}} - \Delta H_{\rm tol} \ . \tag{3.12}$$

In the presence of a heat reservoir at temperature $T$, the information $\Delta I_{\rm min}$ has an energy cost $k_B T \ln 2 \, \Delta I_{\rm min}$ on erasure, which should be compared to the gain in extractable work due to the observation, $k_B T \ln 2 \, (\Delta H_{\mathcal{S}} - \Delta H_{\rm tol})$.

We are now in a position to define hypersensitivity to perturbation. We say a system is *hypersensitive to perturbation* if, for almost all values of $\Delta H_{\rm tol}$, the information $\Delta I_{\rm min}$ is large compared with the corresponding entropy reduction $\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}$, i.e.,

$$\frac{\Delta I_{\rm min}}{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}} \gg 1 \ . \tag{3.13}$$

In terms of energy this definition says that, for a system displaying hypersensitivity to perturbation, possible gains in system free energy through observations of the environment are negligible compared to the Landauer erasure cost of the observational records.

Hypersensitivity to perturbation requires that the inequality (3.13) hold for almost all values of $\Delta H_{\rm tol}$. The inequality (3.13) tends always to hold for sufficiently small values of $\Delta H_{\rm tol}$. The reason is that for these small values of $\Delta H_{\rm tol}$, one is gathering enough information from the perturbing environment to track a particular system state whose entropy is nearly equal to the initial system entropy $H_0$. In other words, one is essentially tracking a particular realization of the perturbation among all possible realizations. Thus, for small values of $\Delta H_{\rm tol}$, the information $\Delta I_{\rm min}$ is a property of the perturbation, being the information to specify a particular realization of the perturbation. The important regime for assessing hypersensitivity to perturbation is thus where $\Delta H_{\rm tol}$ is near to $\Delta H_{\mathcal{S}}$, and it is

in this regime that one can hope that $\Delta I_{\min}$ reveals something about the system dynamics, rather than properties of the perturbation.

In earlier publications [8,10–12], we have conjectured that chaotic Hamiltonian systems, classical or quantum, show hypersensitivity to perturbation. For classical chaotic systems, this can be made plausible in the following way. Under chaotic time evolution, the Liouville density develops structure on finer and finer scales. This highly structured pattern is not itself complex in the algorithmic sense—it is completely specified by the initial density, the Hamiltonian, and the elapsed time—but it can be perturbed in an enormous number of ways [8]. This means that the unperturbed pattern lies very close to a large number of highly complex patterns and that the information about the perturbation needed to specify the perturbed pattern can be very large. In Sec. V we go beyond this heuristic argument and give a proof that a large class of classical chaotic Hamiltonian systems exhibit an *exponential* hypersensitivity to perturbation, in which the ratio (3.13) of information to entropy reduction grows exponentially with time, with the exponential rate of growth given by the KS entropy of the chaotic dynamics. We find that for this class of chaotic systems, the exponential hypersensitivity to perturbation is to a large extent independent of the exact nature of the perturbations and, in particular, of the strength of the perturbations.

In the following sections we limit our investigation to discrete maps. There are two natural ways in which a Hamiltonian flow $\phi_t : X \to X$ on the phase space $X$ induces a discrete map. For an arbitrary time step $\tau$, a map $f : X \to X$ is defined by $fx = \phi_\tau x$ for all $x \in X$. Since $\phi_t \phi_s x = \phi_{t+s} x$ for all times $t$ and $s$ and all $x \in X$, the map $f$ and the flow $\phi_t$ are closely related by $f^n x = \phi_{n\tau} x$ for all $x \in X$ and all integer $n$. Alternatively, a discrete map can be defined via a Poincaré surface of section [1]. The stochastic perturbation of the flow induces a stochastic perturbation of the map at each step.

## IV. SYMBOLIC DYNAMICS

The basic idea underlying the method of symbolic dynamics is to simplify the analysis of dynamical systems by representing points in phase space by symbolic sequences. Parts of the following discussion closely follow [2].

A *discrete abstract dynamical system* $(M, \mu, f)$ consists of a measurable space $M$ with a normalized measure $\mu$ and a measure-preserving automorphism $f$ on $M$, i.e., $\mu(M) = 1$ and $\mu(fA) = \mu(A)$ for all measurable $A$ [26,27]. A *measurable partition* $\mathcal{E}$ of $M$ is defined as a collection $\mathcal{E} = \{E_1, \ldots, E_m\}$ of measurable sets such that

$$\bigcup_{i=1}^{m} E_i = M \quad \text{and} \quad \sum_{i=1}^{m} \mu(E_i) = 1 \; . \tag{4.1}$$

Consider an $m$-letter alphabet $\mathcal{L} = \{1, \ldots, m\}$ where each letter corresponds to one of the $m$ sets in the partition $\mathcal{E}$. We denote by $\omega = \cdots \omega_{-1} \omega_0 \omega_1 \omega_2 \cdots$ a bi-infinite sequence of letters $\omega_n \in \mathcal{L}$ and by $\Sigma$ the set of all such *symbolic sequences*.

For each $x \in M$ we define the set $\Sigma_x \subseteq \Sigma$ as follows:

$$\Sigma_x \equiv \left\{ \omega \;\middle|\; x \in \bigcap_{n=-\infty}^{\infty} f^{-n} E_{\omega_n} \right\} . \tag{4.2}$$

Equivalently, one can say that $\omega \in \Sigma_x \iff f^n x \in E_{\omega_n}$ for all $n$. The set

$$\Sigma_{\mathcal{E}} = \bigcup_{x \in M} \Sigma_x \subseteq \Sigma \tag{4.3}$$

of all symbolic sequences corresponding to at least one point in $M$ is called the set of *admissible sequences*. The partition $\mathcal{E}$ is called a *generating partition* if for each $\omega \in \Sigma_{\mathcal{E}}$ the intersection

$$\bigcap_{n=-\infty}^{\infty} f^{-n} E_{\omega_n} \tag{4.4}$$

consists of only one point, i.e., if each admissible symbolic sequence defines a unique point in $M$. In general, even for generating partitions, $\Sigma_x$ may have more than one element, which means that a point $x \in M$ may be represented by several symbolic sequences $\omega \in \Sigma_x$. For a generating partition, the picture one should have is that the set $\Sigma_{\mathcal{E}}$ of all admissible sequences is the union of disjoint subsets $\Sigma_x$, which may have more than one member.

Let us further define *symbolic words* as finite symbolic sequences $\omega_{n_1} \cdots \omega_{n_2}$ where $n_1 \leq n_2$. In analogy with Eq. (4.4), we define the set of points corresponding to the symbolic word $\omega_{n_1} \cdots \omega_{n_2}$ by

$$E_{\omega_{n_1} \cdots \omega_{n_2}} = \bigcap_{n=n_1}^{n_2} f^{-n} E_{\omega_n} \ . \tag{4.5}$$

We denote by $\Sigma^{(n_1,n_2)}$ the set of all symbolic words $\omega_{n_1} \cdots \omega_{n_2}$. The symbolic word $\omega_{n_1} \cdots \omega_{n_2}$ is *admissible* if $E_{\omega_{n_1} \cdots \omega_{n_2}}$ contains at least one point; we denote by $\Sigma_{\mathcal{E}}^{(n_1,n_2)}$ the set of admissible symbolic words $\omega_{n_1} \cdots \omega_{n_2}$. The $N$th refinement $\mathcal{E}^N$ of the partition $\mathcal{E}$, defined by

$$\mathcal{E}^N = \{E_{\omega_0 \cdots \omega_{N-1}} \mid \omega_0 \cdots \omega_{N-1} \text{ admissible}\} \ , \tag{4.6}$$

is also a measurable partition. If $\mathcal{E}$ is a generating partition, then all refinements of $\mathcal{E}$ are also generating partitions. Furthermore, if $\mathcal{E}$ is generating, then the sigma algebra generated by all refinements of $\mathcal{E}$ coincides with the sigma algebra of all measurable subsets of $M$ [28–30]. The measure $\mu$ induces a measure on the sigma algebra generated by the set of all symbolic words via

$$\mu(\omega_{n_1} \cdots \omega_{n_2}) = \mu(E_{\omega_{n_1} \cdots \omega_{n_2}}) \ . \tag{4.7}$$

Let us also define a conditional measure

$$\mu(\omega_{n_1} \cdots \omega_{n_2} | \omega_{n_2+1} \omega_{n_2+2} \cdots) = \lim_{n \to \infty} \frac{\mu(\omega_{n_1} \cdots \omega_{n_2+n})}{\mu(\omega_{n_2+1} \cdots \omega_{n_2+n})} \tag{4.8}$$

whenever the limit on the right-hand side exists, which is the case for $K$ systems (see below).

The *entropy* $H(\mathcal{E}^N)$ of the refinement $\mathcal{E}^N$ is defined by

$$H(\mathcal{E}^N) = -\sum_{E_{\omega_0 \cdots \omega_{N-1}} \in \mathcal{E}^N} \mu(E_{\omega_0 \cdots \omega_{N-1}}) \log_2 \mu(E_{\omega_0 \cdots \omega_{N-1}}) \ . \tag{4.9}$$

The *metric entropy* or *Kolmogorov-Sinai (KS) entropy* of the map $f$ is defined as

$$h_\mu(f) = \sup_{\mathcal{E}} h_\mu(f|\mathcal{E}) \ , \tag{4.10}$$

where the supremum is taken over all measurable partitions $\mathcal{E}$ and where

$$h_\mu(f|\mathcal{E}) = \lim_{N\to\infty} \frac{H(\mathcal{E}^N)}{N} \ . \tag{4.11}$$

If $\mathcal{E}$ is generating, then $h_\mu(f) = h_\mu(f|\mathcal{E})$ [28–30]. Systems with a positive KS entropy are called $K$ *systems*. Despite its name, the KS entropy is quite different from the Gibbs entropy, for two reasons: (i) $H(\mathcal{E}^N)$ has nothing directly to do with probabilities on the phase space $M$, but is the Shannon information of the ensemble of sets in $\mathcal{E}^N$, when the probability of each set is given by its measure; (ii) $h_\mu(f|\mathcal{E})$ is not an entropy at all, but rather is the asymptotic *rate of increase* of $H(\mathcal{E}^N)$.

A dynamical system is called *ergodic* if time averages equal ensemble averages, i.e., if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \phi(f^n x) = \int_M d\mu\, \phi \quad \text{for almost all } x \in M, \tag{4.12}$$

for any $\mu$-integrable function $\phi$ [26]. All $K$ systems are ergodic [26].

The map $f$ induces a particularly simple so-called *shift map* $\sigma : \Sigma \to \Sigma$ on the set of symbolic sequences. The shift map is defined as

$$(\sigma\omega)_n = \omega_{n+1} \text{ for all } n; \tag{4.13}$$

i.e., $\sigma$ shifts the entire symbolic sequence to the left. The shift map can be extended to a map $\sigma : \Sigma^{(n_1,n_2)} \to \Sigma^{(n_1-1,n_2-1)}$ that acts on symbolic words $\omega_{n_1}\cdots\omega_{n_2} \in \Sigma^{(n_1,n_2)}$ via

$$[\sigma(\omega_{n_1}\cdots\omega_{n_2})]_n = \omega_{n+1} \text{ for } n_1 - 1 < n < n_2 - 1. \tag{4.14}$$

The set of admissible sequences is invariant under the shift map, i.e.,

$$\sigma(\Sigma_\mathcal{E}) = \Sigma_\mathcal{E} \ . \tag{4.15}$$

Furthermore, for a generating partition $\mathcal{E}$, the map $\pi : \Sigma_\mathcal{E} \to M$ defined by

$$\pi(\omega) = \bigcap_{n=-\infty}^{\infty} f^{-n} E_{\omega_n} \tag{4.16}$$

[i.e., $\pi(\omega) = x \iff \omega \in \Sigma_x$] is single-valued and continuous [2]. If the sets $E_i$ forming the partition $\mathcal{E}$ are not mutually exclusive, then the map $\pi$ is not one-to-one. The overlap between different sets $E_i$, however, is of measure zero. The relation between $f$ and $\sigma$ can be summarized in the following commutation diagram:

$$\begin{array}{ccc} \Sigma_\mathcal{E} & \xrightarrow{\sigma} & \Sigma_\mathcal{E} \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ M & \xrightarrow{f} & M \end{array} \ . \tag{4.17}$$

The action of $f$ on measurable subsets of $M$ is faithfully represented by the action of $\sigma$ on measurable sets of symbolic sequences. In the following section, we use this representation to study hypersensitivity to perturbation for $K$ systems.

For the remainder of this section, we assume that $f$ is a $K$ system with KS entropy $h$ and that $\mathcal{E}$ is a generating partition. Since the set $\Sigma_{\mathcal{E}}$ of admissible symbolic sequences is invariant under the action of the shift map $\sigma$ according to Eq. (4.15), $\Sigma_{\mathcal{E}}$ is a stationary source in the language of information theory [31]. Moreover, by choosing the function $\phi$ in Eq. (4.12), for an arbitrary symbolic word $\tilde{\omega} = \omega_{n_1} \dots \omega_{n_2}$, to be the indicator function of the set $E_{\tilde{\omega}}$ [see Eq. (4.5)] corresponding to $\tilde{\omega}$, i.e.,

$$\phi_{\tilde{\omega}}(x) = \begin{cases} 1 \ , & \text{if } x \in E_{\tilde{\omega}}, \\ 0 \ , & \text{otherwise,} \end{cases} \tag{4.18}$$

one sees that $\Sigma_{\mathcal{E}}$ is an ergodic source since $f$ is ergodic.

According to the Shannon-McMillan theorem, stationary ergodic sources have the *asymptotic equipartition property* [31]. This means crudely that for sufficiently large $n$ and arbitrary $n_1$, the set $\Sigma_{\mathcal{E}}^{(n_1, n_1+n-1)}$ of admissible symbolic words of length $n$ consists of approximately $2^{nh}$ symbolic words, each approximately of measure $2^{-nh}$, whereas each of the remaining symbolic words has negligible measure. The choice of $n_1$ is irrelevant because the source is stationary. Formally, a source has the asymptotic equipartition property if and only if for any $\epsilon > 0$ there is a positive integer $n_0(\epsilon)$ such that, for $n > n_0(\epsilon)$ and arbitrary $n_1$, the set $\Sigma_{\mathcal{E}}^{(n_1, n_1+n-1)}$ of admissible symbolic words of length $n$ decomposes into two sets $\Pi$ and $T$ satisfying

$$\sum_{\tilde{\omega} \in \Pi} \mu(\tilde{\omega}) < \epsilon \tag{4.19}$$

and

$$2^{-n(h+\epsilon)} < \mu(\tilde{\omega}) < 2^{-n(h-\epsilon)} \ \text{ for all } \tilde{\omega} \in T. \tag{4.20}$$

## V. PERTURBED CHAOTIC MAPS

Let $(M, \mu, f)$ be a discrete abstract dynamical system that is derived from a Hamiltonian phase-space flow as described at the end of Sec. III. This means, in particular, that the measure $\mu$ is the standard phase-space measure, in units such that the accessible volume of phase space is unity. At the $n$th step the effect of the unperturbed system dynamics is to change the phase-space density from the density $\rho(x, n-1)$ that emerges from the $(n-1)$th step to a new density

$$\rho'(x, n) = \rho(f^{-1}x, n-1) \ . \tag{5.1}$$

We model a measure-preserving stochastic perturbation by alternating unperturbed time steps with application of measure-preserving *perturbation maps*. More precisely, we do the following. We have available a collection of measure-preserving perturbation maps. At the $n$th step we select randomly a particular perturbation map $\xi : M \to M$ from this collection

and apply it to the density $\rho'(x,n)$ that is produced by the unperturbed time step. This yields a new density

$$\rho(x,n) = \rho'(\xi^{-1}x,n) = \rho\Big(f^{-1}(\xi^{-1}x), n-1\Big) \ , \tag{5.2}$$

which depends on the map $\xi$ and which is the input to the next step.

We characterize the perturbation maps in terms of two quantities: (i) the "strength" of the perturbation, which is roughly the size of the phase-space displacements produced by the maps, and (ii) the "correlation cells," which are roughly the phase-space regions over which the displacements produced by the maps remain correlated. We pause here to give a more precise general definition of perturbation strength, because it highlights an essential feature of chaotic dynamics. We defer defining the concept of correlation cells precisely till it emerges naturally in the context of the symbolic dynamics of perturbed chaotic maps. We return to both concepts in Sec. VI, where they are used to develop a heuristic picture of hypersensitivity to perturbation.

To characterize the "strength" of a perturbation, we let $\delta(x_1, x_2)$ denote the Euclidean distance between the two points $x_1, x_2 \in M$ relative to some fixed set of canonical coördinates. An *$\epsilon$-perturbation map* is a perturbation map $\xi$ for which $\delta(\xi x, x) < \epsilon$ for all $x \in M$. An $\epsilon$-perturbation map describes a perturbation whose strength is smaller than the scale set by $\epsilon$.

Now suppose that the initial density $\rho(x, n=0)$ is well behaved in the sense that there is a scale on which $\rho(x, n=0)$ varies little; i.e., there is an $\epsilon_0 > 0$ such that $\rho(x_1, n=0) \simeq \rho(x_2, n=0)$ for any pair of points $x_1, x_2 \in M$ with $\delta(x_1, x_2) < \epsilon_0$. Then, for any integer $n > 0$, there is an $\epsilon > 0$ such that $\rho(x,n)$ varies little on the scale of $\epsilon$. We say that the system is *effectively shielded against perturbations* at the $n$th step if there is an $\epsilon > 0$ such that the perturbation is described by $\epsilon$-perturbation maps and the density $\rho(x,n)$ varies little on the scale of $\epsilon$.

One of the defining properties of chaotic dynamics is that the scale $\epsilon$ on which the density varies little decreases exponentially with the number of time steps $n$. This entails that chaotic systems cannot be effectively shielded against perturbations, except for a small number of time steps. We use this fact below as the starting point for developing an essentially universal description of perturbed chaotic dynamics. Regular, or integrable systems have no exponential relationship between $\epsilon$ and $n$ and thus cannot be fitted within the analysis of this section. We thus defer discussion of regular systems until we have developed a heuristic picture of hypersensitivity to perturbation in Sec. VI.

We now proceed to show that all $K$ systems for which there is a generating partition exhibit hypersensitivity to perturbation. This includes all $K$ systems that have a Markov partition [2]. Assume that the discrete abstract dynamical system $(M, \mu, f)$ has a finite KS entropy $h = h_\mu(f) > 0$, and let $\mathcal{E} = \{E_1, \ldots E_m\}$ be a generating partition of $M$. As explained in Sec. IV, $f$ can be represented by a shift map $\sigma$ on the set of admissible symbolic sequences $\Sigma_\mathcal{E}$, each admissible symbolic sequence corresponding to a single point in $M$. In the following, we identify symbolic sequences with the corresponding points and symbolic words with the corresponding subsets of $M$, writing, e.g., "the symbolic word $\omega_{n_1} \cdots \omega_{n_2}$" when we really mean the set of points corresponding to the symbolic word $\omega_{n_1} \cdots \omega_{n_2}$. The set of admissible symbolic sequences has the asymptotic equipartition property; i.e., for $n \gg 1$, $M$ is partitioned by the admissible symbolic words $\omega_1 \cdots \omega_n$ in such a way that there

are approximately $2^{nh}$ symbolic words each approximately of measure $2^{-nh}$, whereas each of the remaining symbolic words has negligible measure.

Let us first look at the unperturbed evolution of a simple initial state on $M$. We assume that the initial density $\rho(x, n = 0)$ is constant on the set of points corresponding to the symbolic word

$$\Omega = \omega_{n_0+1} \cdots \omega_{n_0+q} \ , \quad q \gg 1, \tag{5.3}$$

and zero elsewhere. Here $\omega_{n_0+1} \cdots \omega_{n_0+q}$ is one of the symbolic words that has measure

$$\mu_0 \equiv \mu(\omega_{n_0+1} \cdots \omega_{n_0+q}) \simeq 2^{-qh} \ . \tag{5.4}$$

In the following, we refer to a subset of $M$ on which the density is constant as a *pattern*.

We choose the (arbitrary) zero of the entropy such that the entropy of a uniform density constant on the entire set $M$ vanishes. This is a natural choice because it corresponds to choosing units such that $\mu(x)$ is the measure in the Gibbs entropy (2.1). The entropy of the initial density is thus

$$H_0 = \log_2 \mu_0 \simeq \log_2(2^{-qh}) = -qh \ . \tag{5.5}$$

The condition $q \gg 1$ means that the initial entropy $H_0$ is much smaller than the negative of the KS entropy of the map, $-h$.

Applying the shift map $\sigma$ for $n$ steps leads to a uniform density on $\Omega' = \sigma^n \Omega = \omega'_{n_0+1-n} \cdots \omega'_{n_0+q-n}$ where $\omega'_k = \omega_{k+n}$. The entropy of the shifted pattern remains unchanged. As was stressed in Sec. II, the entropy does not change under unperturbed Hamiltonian evolution. Moreover, the method of symbolic dynamics makes it utterly obvious that *no additional information beyond the initial pattern and the number of steps n is needed to give a complete description of the evolved pattern.* As was pointed out in Sec. III, the evolved unperturbed density, though highly structured when viewed in phase space, is not complex in the algorithmic sense.

We now turn to perturbed evolution. At each step, instead of applying just the map $f$, we now apply first $f$ and then a measure-preserving map $\xi$ selected randomly from our collection of maps. We make two major assumptions about the perturbation maps $\xi$ and their probabilities, the first assumption having to do with the perturbation strength and the second with the perturbation correlation cells. The first assumption is that below some scale on phase space, a single application of the perturbation randomizes the pattern completely. In symbolic language this scale is characterized by some negative integer $-n_p$, and our assumption can be written as

$$\begin{aligned} \mathrm{Prob}\Big((\xi\omega)_k = \omega'_k, \ k = n, \ldots, -n_p\Big) &= \mu(\omega'_n \cdots \omega'_{-n_p} |\, \omega_{-n_p+1}\omega_{-n_p+2} \cdots) \\ &\text{for all } n \le -n_p, \ \omega \in \Sigma_{\mathcal{E}}, \text{ and } \omega'_n \cdots \omega'_{-n_p} \in \Sigma_{\mathcal{E}}^{(n,-n_p)} \ , \end{aligned} \tag{5.6}$$

where Prob stands for probability with respect to the random selection of the perturbation map. The integer $n_p$ is a measure of the strength of the perturbation, large $n_p$ meaning a weak perturbation. Another way of describing this assumption is the following: take a point on phase space—i.e., a symbolic sequence $\omega$—and perturb it to get a new point $\xi\omega$; Eq. (5.6) means that it is unpredictable, relative to the random selection of perturbation map $\xi$, in

which partition element $E_i$ the $n_p$th backward iterate and all further backward iterates of the perturbed point $\xi\omega$ fall.

Our second major assumption concerns the perturbation correlation cells. We assume that the perturbation maps $\xi$ are not completely arbitrary, but that a particular map displaces neighboring points in a similar way and that, averaged over the random selection of perturbation maps, the displacements become uncorrelated for points sufficiently far away from each other. We model this behavior by assuming that the space $M$ is partitioned into *perturbation cells* $\omega_{-n_p+s+1}\cdots\omega_{-n_p+s+r}$, where $r \gg 1$ and $s \geq 0$ are integers, such that, first, the perturbations are uncorrelated for points in different perturbation cells and, second, knowing how a typical point in a perturbation cell is perturbed determines how all points in that cell are perturbed. These perturbation cells are a precise realization of the notion of correlation cells.

In addition to our two major assumptions, we make several simplifying assumptions or approximations about the perturbation maps. These simplifying assumptions always tend to reduce the information $\Delta I_{\min}$ required to reduce the entropy increase to the tolerable amount $\Delta H_{\text{tol}}$. Since we want to prove that $\Delta I_{\min}$ is large, such simplifying assumptions do not limit the validity of our results. As our first simplifying assumption, we ignore all features of the perturbation maps beyond what is needed to satisfy Eq. (5.6); i.e., we choose perturbation maps $\xi$ that satisfy

$$(\xi\omega)_n = \omega_n \text{ for all } n > -n_p \text{ and } \omega \in \Sigma_{\mathcal{E}}, \tag{5.7}$$

in addition to Eq. (5.6). This assumption means that the perturbation maps have no effect at all on scales larger than the scale set by $n_p$. Allowing the perturbation maps to act on scales larger than that set by $n_p$ would lead to more distinguishable perturbed patterns—and thus to higher $\Delta I_{\min}$—which would have to be tracked to keep the entropy increase to some tolerable amount.

Since it is impossible to shield a chaotic system against perturbations in the sense defined above, we are justified in choosing the zero of time ($n = 0$) such that the perturbation becomes effective at the first time step ($n = 1$). This amounts to choosing the initial symbolic word (5.3) so that $n_0 = -n_p$, where $n_p$ is the integer that characterizes the strength of the perturbation. This initial symbolic word, which defines the pattern on which the initial density $\rho(x, n = 0)$ is nonzero, can thus be written as

$$\Omega = \mid \omega_{-n_p+1}\cdots\omega_{-n_p+s} \mid \omega_{-n_p+s+1}\cdots\omega_{-n_p+s+r} \mid \cdots\omega_{-n_p+q} \, . \tag{5.8}$$

Since the perturbation maps satisfy Eq. (5.7), the perturbation leaves the pattern of Eq. (5.8) unchanged. After one time step, however, the leftmost symbol moves into the *perturbation region*, located to the left of the leftmost vertical bar in Eq. (5.8), where it is randomized by the perturbation according to Eq. (5.6). The perturbation region is separated by $s$ letters from the *decision region*, located between the middle and rightmost vertical bars in Eq. (5.8). This decision region, $r$ letters wide, defines the perturbation cells. Since we assume that $r \gg 1$, there are approximately $2^{rh}$ *typical perturbation cells*, each of size $\simeq 2^{-rh}$, whereas the total size of the remaining perturbation cells can be neglected. Even though the assumption $q > r + s$ is implicit in the way we write the initial word in Eq. (5.8), this assumption is not necessary for our analysis.

Focus attention now on the phase-space density $\rho(x,n)$ after $n$ time steps, where we assume that

$$q - s \geq n \gg \max\Big(1, q - s - r\Big) \; . \tag{5.9}$$

These assumptions assure us that the leftmost letter of the initial word (5.8) has moved deep into the perturbation region and the rightmost letter has moved far to the left of the right boundary of the decision region, but not more than one position beyond the left boundary of the decision region. After $n$ *unperturbed* steps the initial pattern $\Omega$ given by Eq. (5.8) evolves into the pattern $\Omega' = \sigma^n \Omega$, which has the form

$$\Omega' = \omega'_{-n_p-n+1} \cdots \omega'_{-n_p} \mid \omega'_{-n_p+1} \cdots \omega'_{-n_p+s} \mid \omega'_{-n_p+s+1} \cdots \omega'_{-n_p-n+q} \; , \tag{5.10}$$

where $\omega'_k = \omega_{k+n}$.

Consider now what happens when the pattern of Eq. (5.10) is perturbed. According to Eq. (5.6), all $n$ letters in the perturbation region [to the left of the leftmost vertical bar in Eq. (5.10)] are randomized by the perturbation. We can therefore ignore the effect of perturbations applied at previous steps. The density that arises from averaging over the perturbation is made up of all the patterns that come from randomizing the letters in the perturbation region. As a consequence of the asymptotic equipartition property and assumption (5.6), there are approximately $2^{nh}$ such patterns, all of which have approximately the same probability and all of which have approximately the same measure as the unperturbed pattern (5.10). Thus averaging over the perturbation leads to an entropy increase

$$\Delta H_{\mathcal{S}} \simeq \log_2(2^{nh}) = nh \; . \tag{5.11}$$

We now turn to estimating the minimum information $\Delta I_{\min}$ about the perturbation needed to limit the entropy increase to a tolerable value $\Delta H_{\rm tol}$. Consider again the word (5.10) that describes the unperturbed pattern after $n$ steps. Due to the asymptotic equipartition property, the $n - (q - s - r) \gg 1$ unspecified letters at the right side of the decision region correspond to the pattern's extending over

$$\mathcal{R}_n \equiv 2^{[n-(q-s-r)]h} \gg 1 \tag{5.12}$$

typical perturbation cells. This exponential increase in the number of typical perturbation cells occupied by the pattern continues only until all the typical perturbation cells are occupied, i.e., until $\mathcal{R}_n = 2^{rh}$ or $n = q - s$. The occupied perturbation cells partition the unperturbed pattern into $\mathcal{R}_n$ *sub-patterns* of the form

$$\omega'_{-n_p-n+1} \cdots \omega'_{-n_p} \mid \omega'_{-n_p+1} \cdots \omega'_{-n_p+s} \mid \omega'_{-n_p+s+1} \cdots \omega'_{-n_p-n+q} \hat{\omega}_{-n_p-n+q+1} \cdots \hat{\omega}_{-n_p+s+r} \mid \; , \tag{5.13}$$

where the $n - (q - s - r)$ letters $\hat{\omega}_i$ determine an occupied perturbation cell.

These sub-patterns, all of approximately the same size, are perturbed independently. We describe the *perturbed* sub-pattern in each perturbation cell by a symbolic word

$$\tilde{\omega}_{-n_p-n+1} \cdots \tilde{\omega}_{-n_p} \mid \omega'_{-n_p+1} \cdots \omega'_{-n_p+s} \mid \omega'_{-n_p+s+1} \cdots \omega'_{-n_p-n+q} \hat{\omega}_{-n_p-n+q+1} \cdots \hat{\omega}_{-n_p+s+r} \mid \; , \tag{5.14}$$

where the letters $\tilde{\omega}_i$ are chosen at random according to Eq. (5.6). Again invoking the asymptotic equipartition property, we can say that in each of the $\mathcal{R}_n$ occupied perturbation cells, there are

$$D \equiv 2^{nh} \gg 1 \tag{5.15}$$

typical perturbed words, or typical perturbed sub-patterns, of the form (5.14), all having approximately the same probability and all having approximately the same measure as the unperturbed sub-pattern (5.13).

These considerations give a total of $D^{\mathcal{R}_n}$ typical perturbed patterns, all produced with approximately the same probability by the perturbation and all having approximately the same entropy as the unperturbed pattern (5.10). The information needed to specify a particular perturbed pattern—and thus the information needed to keep the tolerable entropy increase essentially to zero—is given by

$$\Delta I_{\min} \simeq \mathcal{R}_n \log_2 D \simeq 2^{[n-(q-s-r)]h} \Delta H_{\mathcal{S}} \quad \text{for } \Delta H_{\text{tol}} \simeq 0. \tag{5.16}$$

It should be emphasized that the exponential increase of this $\Delta I_{\min}$ continues only until all the typical perturbation cells are occupied, i.e., until $n = q - s$; for $n > q - s$ the information continues to increase, but the form of the increase is more difficult to determine.

What is going on here has a simple interpretation. Within each perturbation cell, the perturbed sub-patterns have essentially no overlap. The overall perturbed patterns, however, can have considerable overlap, since two perturbed patterns are different even if they differ in only a single perturbation cell. The entropy increase $\Delta H_{\mathcal{S}} \simeq nh$ that comes from averaging over the perturbation [Eq. (5.11)] is the logarithm of the number $D$ of *non-overlapping* patterns that are required to make up the average density. The number of non-overlapping patterns is the same as the number of perturbed sub-patterns in each perturbation cell, and hence $\Delta H_{\mathcal{S}} \simeq nh$ is also the information required to specify a particular sub-pattern within a perturbation cell. To specify a particular overall pattern, however, one must say which perturbed sub-pattern is realized in each of the $\mathcal{R}_n$ occupied perturbation cells; this requires giving $\Delta H_{\mathcal{S}} \simeq nh$ bits per occupied perturbation cell, for a total amount of information $\Delta I_{\min} \simeq \mathcal{R}_n \Delta H_{\mathcal{S}}$ [Eq. (5.16)]. The information $\Delta I_{\min}$ is much bigger than the average entropy increase $\Delta H_{\mathcal{S}}$ because the information counts overlapping patterns, whereas the entropy does not.

Now suppose that one allows a nonzero tolerable entropy increase $\Delta H_{\text{tol}}$. This means that one does not have to specify exactly which of the $D^{\mathcal{R}_n}$ perturbed patterns is realized. Instead, one can group the typical perturbed patterns and specify only to which group the perturbed pattern belongs. Suppose the typical patterns are grouped into $N$ groups, which are labeled by an integer $b = 1, \ldots, N$. In analogy to Sec. III, we denote by $N_b$ the number of patterns in the $b$th group ($\sum_{b=1}^{N} N_b = D^{\mathcal{R}_n}$), by $\rho_b(x)$ the probability density one obtains by averaging over all the patterns in the $b$th group, by $\Delta H_b$ the corresponding conditional entropy increase, and by $\overline{\Delta H} = \sum p_b \Delta H_b$ the average conditional entropy increase. Since all the patterns are approximately equi-probable, the probability of obtaining the measurement record $b$, which specifies that the perturbed pattern is in the $b$th group, is $p_b = N_b D^{-\mathcal{R}_n}$.

To obtain $\Delta I_{\min}$ for a given $\Delta H_{\text{tol}}$, one would have to find a grouping of the patterns that is optimal in the sense of minimizing $\Delta I_{\min}$ under the condition that $\overline{\Delta H} \leq \Delta H_{\text{tol}}$. Since we do not know how to find an optimal grouping, we construct a nearly optimal grouping as

follows. We start with a particular pattern, or *fiducial pattern*, and form our first group out of all the patterns that differ in at most $d$ perturbation cells from the fiducial pattern. Such a group we call a $d$-group. The grouping into $d$-groups is motivated by the fact that the entropy increase $\overline{\Delta H}$ is minimal for groups of patterns that differ in the smallest number of perturbation cells [see Eq. (5.27) below]. There being

$$g_k = \binom{\mathcal{R}_n}{k}(D-1)^k \tag{5.17}$$

patterns that differ in exactly $k$ cells from an arbitrary fiducial pattern, the number of patterns differing in at most $d$ cells from an arbitrary fiducial pattern and therefore the size of a $d$-group is

$$G_d = \sum_{k=0}^{d} g_k = \sum_{k=0}^{d} \binom{\mathcal{R}_n}{k}(D-1)^k \; . \tag{5.18}$$

A particularly simple way to proceed would be to pick a second fiducial pattern from among the patterns not in the first group, forming a second $d$-group about this second pattern, and then to continue to form $d$-groups until all patterns were grouped. Unfortunately, this strategy fails because if we proceed in this way, some groups overlap. The problem of finding a grouping into non-overlapping $d$-groups is equivalent to the problem of finding a perfect error-correcting code in information theory [31] and generally has no solution. In the following, we nevertheless assume that the $D^{\mathcal{R}_n}$ patterns are perfectly grouped into a number $N = D^{\mathcal{R}_n}/G_d$ of $d$-groups. We can make this simplifying assumption because it lowers our estimate of $\Delta I_{\min}$.

We now turn to the computation of the entropy increase $\Delta H_d$ for a $d$-group, i.e., a group consisting of a fiducial pattern and all the patterns differing in at most $d$ perturbation cells from the fiducial pattern. The average density $\rho_d(x)$ for a $d$-group is the average of the densities for the $G_d$ patterns in the group, all patterns contributing with the same probability $1/G_d$. Alternatively, we can break each contributing pattern into its $\mathcal{R}_n$ sub-patterns—i.e., symbolic words of the form (5.14)—and view $\rho_d(x)$ as being made up of contributions from the $D\mathcal{R}_n$ sub-patterns, all of which have approximately the same measure $\mu_0/\mathcal{R}_n$.

We distinguish two types of sub-patterns, namely the $\mathcal{R}_n$ sub-patterns belonging to the fiducial pattern and the other $(D-1)\mathcal{R}_n$ sub-patterns. The average density $\rho_d(x)$ is uniform on each sub-pattern. We denote its value on sub-patterns belonging to the fiducial sub-pattern by $\rho_{df}$ and its value on the other sub-patterns by $\rho_{do}$. For a sub-pattern belonging to the fiducial pattern, the probability obtained by integrating $\rho_d$ over the sub-pattern is

$$\int d\mu(x)\, \rho_d = \rho_{df} \frac{\mu_0}{\mathcal{R}_n} = \frac{p_f}{\mathcal{R}_n} \; , \tag{5.19}$$

where $p_f$ is the probability obtained by integrating $\rho_d$ over the entire fiducial pattern. Similarly, for any of the other sub-patterns, the probability obtained by integrating $\rho_d$ over the sub-pattern is

$$\int d\mu(x)\, \rho_d = \rho_{do} \frac{\mu_0}{\mathcal{R}_n} = \frac{p_o}{(D-1)\mathcal{R}_n} \; , \tag{5.20}$$

where $p_o = 1 - p_f$ is the probability obtained by integrating $\rho_d$ over all the sub-patterns outside the fiducial pattern.

The entropy increase of a $d$-group can now be written as

$$\begin{aligned}\Delta H_d &= -\int d\mu(x)\,\rho_d(x)\log_2[\rho_d(x)] \;- H_0 \\ &= -\mathcal{R}_n\frac{p_f}{\mathcal{R}_n}\log_2\left(\frac{p_f}{\mu_0}\right) - (D-1)\mathcal{R}_n\frac{p_o}{(D-1)\mathcal{R}_n}\log_2\left(\frac{p_o}{\mu_0(D-1)}\right) - \log_2\mu_0 \\ &= -p_f\log_2 p_f - p_o\log_2 p_o + p_o\log_2(D-1)\;. \end{aligned} \tag{5.21}$$

To evaluate $\Delta H_d$, we must find the integrated probabilities $p_o$ and $p_f$. Each pattern that differs in exactly $k$ cells from the fiducial pattern contributes the amount $k/\mathcal{R}_n G_d$ to $p_o$ and the amount $(\mathcal{R}_n - k)/\mathcal{R}_n G_d$ to $p_f$. It follows that

$$p_o = \sum_{k=0}^{d}\frac{k}{\mathcal{R}_n G_d}g_k = \frac{1}{\mathcal{R}_n G_d}\sum_{k=0}^{d}\binom{\mathcal{R}_n}{k}(D-1)^k\,k \tag{5.22}$$

and

$$p_f = \sum_{k=0}^{d}\frac{\mathcal{R}_n - k}{\mathcal{R}_n G_d}g_k = \frac{1}{\mathcal{R}_n G_d}\sum_{k=0}^{d}\binom{\mathcal{R}_n}{k}(D-1)^k\,(\mathcal{R}_n - k)\;. \tag{5.23}$$

Notice that $p_o = \mathcal{R}_n^{-1}\partial\ln G_d/\partial\ln(D-1)$ and that when $d = \mathcal{R}_n$, we have $G_{\mathcal{R}_n} = D^{\mathcal{R}_n}$, $p_o = 1 - 1/D$, $p_f = 1/D$, and thus $\Delta H_{\mathcal{R}_n} = \log_2 D = \Delta H_{\mathcal{S}}$.

Under the assumption of perfect grouping into $N = D^{\mathcal{R}_n}/G_d$ $d$-groups, the average entropy of the $d$-groups is $\overline{\Delta H} = \Delta H_d$, and the information to specify a particular $d$-group is

$$\Delta I_d \simeq \log_2 N \simeq \mathcal{R}_n\log_2 D - \log_2 G_d = \mathcal{R}_n\Delta H_{\mathcal{S}} - \log_2 G_d\;. \tag{5.24}$$

Under our further simplifying assumption that optimal grouping is well approximated by perfect grouping into $d$-groups, we can approximate the minimum information $\Delta I_{\min}$ required to keep the entropy increase to a tolerable amount $\Delta H_{\rm tol}$ by

$$\Delta I_{\min} \simeq \Delta I_d \;\text{ for }\Delta H_{\rm tol} \simeq \Delta H_d\;. \tag{5.25}$$

At this point we could plot $\Delta I_{\min}$ as a function of $\Delta H_{\rm tol}$ by using the common dependence on $d$. Given the assumptions (5.12) and (5.15) that $\mathcal{R}_n$ and $D$ are large, however, we can introduce further approximations that allow us to write an explicit expression for $\Delta I_{\min}$ as a function of $\Delta H_{\rm tol}$, valid over nearly the entire range of $\Delta H_{\rm tol}$. The key to these approximations is that $g_k$ increases exponentially for $k \ll k_c \equiv (\mathcal{R}_n + 1)(1 - 1/D)$. This means that each of the sums for $G_d$, $p_o$, and $p_f$ can be approximated by its largest term ($k = d$), provided $\mathcal{R}_n - d \gg \mathcal{R}_n - k_c \simeq \mathcal{R}_n/D - 1$. The resulting approximations are

$$G_d \simeq \binom{\mathcal{R}_n}{d}(D-1)^d, \quad p_o \simeq \frac{d}{\mathcal{R}_n}, \quad p_f \simeq \frac{\mathcal{R}_n - d}{\mathcal{R}_n}\;. \tag{5.26}$$

In this approximation the entropy increase of a $d$-group is

$$\Delta H_d \simeq -\frac{\mathcal{R}_n - d}{\mathcal{R}_n} \log_2 \frac{\mathcal{R}_n - d}{\mathcal{R}_n} - \frac{d}{\mathcal{R}_n} \log_2 \frac{d}{\mathcal{R}_n} + \frac{d}{\mathcal{R}_n} \log_2 (D-1) \; . \tag{5.27}$$

Using the same approximation and applying Stirling's formula, one finds that

$$\begin{aligned} \log_2 G_d &\simeq \log_2 \binom{\mathcal{R}_n}{d} + d \log_2 (D-1) \\ &\simeq \mathcal{R}_n \left( -\frac{\mathcal{R}_n - d}{\mathcal{R}_n} \log_2 \frac{\mathcal{R}_n - d}{\mathcal{R}_n} - \frac{d}{\mathcal{R}_n} \log_2 \frac{d}{\mathcal{R}_n} + \frac{d}{\mathcal{R}_n} \log_2 (D-1) \right) \\ &\simeq \mathcal{R}_n \Delta H_d \; . \end{aligned} \tag{5.28}$$

Combining Eqs. (5.24), (5.25), and (5.28) yields

$$\Delta I_{\min} \simeq \Delta I_d \simeq \mathcal{R}_n (\Delta H_{\mathcal{S}} - \Delta H_d) \simeq \mathcal{R}_n (\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}) \; . \tag{5.29}$$

This expression, the key result of this paper, shows that to reduce the entropy of a perturbed chaotic map by an amount $\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}$, one must acquire an amount of information $\Delta I_{\min}$ about the perturbation which is much larger than the contemplated entropy reduction. Indeed, the ratio of information to entropy reduction grows exponentially as $\mathcal{R}_n = 2^{[n-(q-s-r)]h}$ with the number of time steps, the exponential rate of growth being determined by the KS entropy $h$ of the map. This is what we mean by exponential hypersensitivity to perturbation.

We should investigate the validity of the approximations that lead to our key result (5.29). This result agrees with what we have already derived in Eq. (5.16) for $\Delta H_{\rm tol} \simeq 0$. Thus we are mainly interested in knowing where the approximations fail as $\Delta H_{\rm tol}$ approaches $\Delta H_{\mathcal{S}}$. A more careful analysis, which keeps track of the errors introduced by the approximation (5.26) and by the use of Stirling's formula in Eq. (5.28), indicates that we must consider separately two cases: (i) $\mathcal{R}_n \gtrsim D$ $(r+s \gtrsim q)$, i.e., there are more occupied perturbation cells than there are perturbed sub-patterns per cell; (ii) $\mathcal{R}_n \lesssim D$ $(r+s \lesssim q)$, i.e., there are fewer occupied perturbation cells than there are perturbed sub-patterns per cell. In case (i), Eq. (5.29) is valid as long as $\mathcal{R}_n - d \gg \mathcal{R}_n/D \gtrsim 1$, which translates to

$$\mathcal{R}_n \gtrsim D\text{:} \;\; \Delta H_{\mathcal{S}} - \Delta H_{\rm tol} \gg \frac{1}{D} \iff \Delta I_{\min} \gg \frac{\mathcal{R}_n}{D} \gtrsim 1 \; ; \tag{5.30}$$

in case (ii), Eq. (5.29) is valid as long as $\mathcal{R}_n - d \gg \ln(eD/\mathcal{R}_n) \gtrsim 1 \gtrsim \mathcal{R}_n/D$, which translates to

$$\mathcal{R}_n \lesssim D\text{:} \;\; \Delta H_{\mathcal{S}} - \Delta H_{\rm tol} \gg \frac{1}{\mathcal{R}_n} \left( \ln\left(\frac{eD}{\mathcal{R}_n}\right) \right)^2 \iff \Delta I_{\min} \gg \left( \ln\left(\frac{eD}{\mathcal{R}_n}\right) \right)^2 \gtrsim 1 \; . \tag{5.31}$$

These restrictions arise because of approximations made in evaluating $\Delta I_d$ and $\Delta H_d$.

There is a separate question of whether perfect $d$-grouping is a good approximation to optimal grouping. The restrictions contained in Eqs. (5.30) and (5.31) are probably not the most important restrictions on the validity of our key result, because the very idea of perfect $d$-grouping as an approximation to optimal grouping is suspect when $\Delta I_{\min}$ is as small as a few bits. Our hesitancy in defining exponential hypersensitivity to perturbation, where we require the information-to-entropy ratio (3.13) to grow exponentially for "almost all" values of $\Delta H_{\rm tol}$, can be traced to this inability to approximate the optimal grouping when $\Delta H_{\rm tol}$ is

very close to $\Delta H_{\mathcal{S}}$. We are left uncertain about the precise behavior of $\Delta I_{\rm min}$ when $\Delta H_{\rm tol}$ is very close to $\Delta H_{\mathcal{S}}$.

We can interpret our key result by hearkening back to the interpretation given to Eq. (5.16). We first need to describe what it means to to specify the phase-space density at a level of resolution defined by a tolerable entropy increase $\Delta H_{\rm tol}$. To do so, imagine that the sub-patterns within each occupied perturbation cell are aggregated into groups, which we call *coarse-grained sub-patterns*, each group consisting of $2^{\Delta H_{\rm tol}}$ sub-patterns so that there are

$$\mathcal{D} = D/2^{\Delta H_{\rm tol}} \simeq 2^{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}} \tag{5.32}$$

coarse-grained sub-patterns in each occupied perturbation cell. A *coarse-grained pattern* consists of coarse-grained sub-patterns, one for each of the $\mathcal{R}_n$ occupied perturbation cells. Since a coarse-grained pattern has a measure that is approximately $2^{\Delta H_{\rm tol}}$ times as big as a pattern, a coarse-grained pattern represents an entropy increase

$$\log_2(2^{\Delta H_{\rm tol}}) = \Delta H_{\rm tol} \ . \tag{5.33}$$

Thus, specifying the system state at a level of resolution set by $\Delta H_{\rm tol}$ amounts to specifying a particular coarse-grained pattern.

The further entropy increase that results from averaging over the coarse-grained patterns is given approximately by

$$\log_2 \mathcal{D} \simeq \Delta H_{\mathcal{S}} - \Delta H_{\rm tol} \ . \tag{5.34}$$

This entropy increase is the logarithm of the number of *non-overlapping* coarse-grained patterns that are required to make up the density that comes from averaging over the perturbation. This number of non-overlapping coarse-grained patterns is the same as the number of coarse-grained sub-patterns in each perturbation cell, and hence the entropy increase (5.34) is also the information required to specify a particular coarse-grained sub-pattern within a perturbation cell. There being $\mathcal{R}_n$ perturbation cells, the information needed to specify an entire coarse-grained pattern becomes

$$\Delta I_{\rm min} \simeq \mathcal{R}_n(\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}) = \mathcal{R}_n \log_2 \mathcal{D} \ , \tag{5.35}$$

an amount of information that corresponds to a total of $\mathcal{D}^{\mathcal{R}_n}$ coarse-grained patterns, all produced with approximately the same probability by the perturbation.

The exponential hypersensitivity to perturbation that we have demonstrated here for maps with positive KS entropy is an asymptotic property for large times. By spelling out precisely the character of the $n \to \infty$ limit, we can see how exponential hypersensitivity to perturbation provides an alternative definition of the KS entropy. In discussing the limit, it is helpful to have in mind the form (5.8) of the initial symbolic word and the form (5.10) of the unperturbed symbolic word after $n$ time steps. The assumptions (5.9) indicate that as $n$ goes to infinity, we should let $n-(q-r-s)$ go to infinity in the same way as $n$—this allows the limit to explore the long-time exponential growth of $\mathcal{R}_n$—while keeping $q-s-n \geq 0$ constant—this prevents the exponential growth of $\mathcal{R}_n$ from being halted at the time when there is more than one sub-pattern per perturbation cell. Thus an appropriate limit is to let

$n$, $q$, and $r$ go to infinity, while keeping $s$, $q - n \geq s$, and $r - n$ constant. In thinking about how this limit is mapped onto phase space, it is convenient also to let $n_p$ go to infinity while keeping $-n_p + q$ constant; this keeps the rightmost letter of the initial symbolic word in the same place as we take the limit. With this understanding of the limit, we can write

$$\lim_{n\to\infty}\left(\frac{1}{n}\log_2\left(\frac{\Delta I_{\min}}{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}}\right)\right) = \lim_{n\to\infty}\left(\frac{\log_2 \mathcal{R}_n}{n}\right) = h \ . \tag{5.36}$$

In terms of phase space, this long-time limit means that the size of the initial pattern, the size of a typical perturbation cell, and the strength of the perturbation all go to zero at the same rate as $n$ goes to infinity.

## VI. DISCUSSION

The objective of this final section is to extract the important ideas from the symbolic dynamics and to use them to develop a simple, heuristic picture of hypersensitivity to perturbation. Consider a classical system whose dynamics unfolds on a $2F$-dimensional phase space, and suppose that the system is perturbed by a stochastic Hamiltonian whose effect can be described as diffusion on phase space.

Suppose first that the system is globally chaotic with KS entropy $K$. For such a system a phase-space density is stretched and folded by the chaotic dynamics, developing exponentially fine structure as the dynamics proceeds. A simple picture is that the phase-space density stretches exponentially in half the phase-space dimensions and contracts exponentially in the other half of the dimensions.

The perturbation is characterized by a perturbation strength and by correlation cells. We can take the perturbation strength to be the typical distance (e.g., Euclidean distance with respect to some fixed set of canonical coördinates) that a phase-space point diffuses under the perturbation during an $e$-folding time, $F/K \ln 2$, in a typical contracting dimension. The perturbation becomes effective, in the sense described in Sec. V, when the phase-space density has roughly the same size in the contracting dimensions as the perturbation strength. Once the perturbation becomes effective, the effects of the diffusive perturbation and of the further contraction roughly balance one another, leaving the *average* phase-space density with a constant size in the contracting dimensions.

The correlation cells are phase-space cells over which the effects of the perturbation are well correlated and between which the effects of the perturbation are essentially uncorrelated. We assume that all the correlation cells have approximately the same phase-space volume. We can get a rough idea of the effect of the perturbation by regarding the correlation cells as receiving independent perturbations. Moreover, the diffusive effects of the perturbation during an $e$-folding time $F/K \ln 2$ are compressed exponentially during the next such $e$-folding time; this means that once the perturbation becomes effective, the main effects of the perturbation at a particular time are due to the diffusion during the immediately preceding $e$-folding time.

Since a chaotic system cannot be forever shielded from the effects of the perturbation, we can choose the initial time $t = 0$ to be the time at which the perturbation is just becoming effective. We suppose that at $t = 0$ the unperturbed density is spread over $2^{-Kt_0}$ correlation

cells, $t_0$ being the time when the unperturbed density occupies a single correlation cell. The essence of the KS entropy is that for large times $t$ the unperturbed density spreads over

$$\mathcal{R}(t) \sim 2^{K(t-t_0)} \tag{6.1}$$

correlation cells, in each of which it occupies roughly the same phase-space volume. The exponential increase of $\mathcal{R}(t)$ continues until the unperturbed density is spread over essentially all the correlation cells. We can regard the unperturbed density as being made up of *sub-densities*, one in each occupied correlation cell and all having roughly the same phase-space volume.

After $t = 0$, when the perturbation becomes effective, the *average* density continues to spread exponentially in the expanding dimensions. This spreading is not balanced, however, by contraction in the other dimensions, so the phase-space volume occupied by the average density grows as $2^{Kt}$, leading to an entropy increase

$$\Delta H_{\mathcal{S}} \sim \log_2(2^{Kt}) = Kt \ . \tag{6.2}$$

Just as the unperturbed density can be broken up into sub-densities, so the average density can be broken up into *average sub-densities*, one in each occupied correlation cell. Each average sub-density occupies a phase-space volume that is $2^{Kt}$ times as big as the volume occupied by an unperturbed sub-density.

The unperturbed density is embedded within the phase-space volume occupied by the average density and itself occupies a volume that is smaller by a factor of $2^{-Kt}$. We can picture a *perturbed* density crudely by imagining that in each occupied correlation cell the unperturbed sub-density is moved rigidly to some new position within the volume occupied by the average sub-density; the result is a *perturbed sub-density*. A *perturbed density* is made up of perturbed sub-densities, one in each occupied correlation cell. All of the possible perturbed densities are produced by the perturbation with roughly the same probability.

Suppose now that we wish to hold the entropy increase to a tolerable amount $\Delta H_{\rm tol}$. We must first describe what it means to specify the phase-space density at a level of resolution set by a tolerable entropy increase $\Delta H_{\rm tol}$. An approximate description can be obtained in the following way. Take an occupied correlation cell, and divide the volume occupied by the average sub-density into $2^{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}}$ non-overlapping volumes, all of the same size. Aggregate all the perturbed sub-densities that lie predominantly within a particular one of these non-overlapping volumes to produce a *coarse-grained sub-density*. There are $2^{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}}$ coarse-grained sub-densities within each occupied correlation cell, each having a phase-space volume that is bigger than the volume occupied by a perturbed sub-density by a factor of

$$\frac{2^{Kt}}{2^{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}}} = 2^{\Delta H_{\rm tol}} \ . \tag{6.3}$$

A *coarse-grained density* is made up by choosing a coarse-grained sub-density in each occupied correlation cell. A coarse-grained density occupies a phase-space volume that is bigger than the volume occupied by the unperturbed density by the factor $2^{\Delta H_{\rm tol}}$ of Eq. (6.3) and hence represents an entropy increase

$$\log_2(2^{\Delta H_{\rm tol}}) = \Delta H_{\rm tol} \ . \tag{6.4}$$

Thus to specify the phase-space density at a level of resolution set by $\Delta H_{\rm tol}$ means roughly to specify a coarse-grained density. The further entropy increase on averaging over the perturbation is given by

$$\log_2(2^{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}}) = \Delta H_{\mathcal{S}} - \Delta H_{\rm tol} \ . \tag{6.5}$$

What about the information $\Delta I_{\rm min}$ required to hold the entropy increase to $\Delta H_{\rm tol}$? Since there are $2^{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}}$ coarse-grained sub-densities in an occupied correlation cell, each produced with roughly the same probability by the perturbation, it takes approximately $\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}$ bits to specify a particular coarse-grained sub-density. To describe a coarse-grained density, one must specify a coarse-grained sub-density in each of the $\mathcal{R}(t)$ occupied correlation cells. Thus the information required to specify a coarse-grained density—and, hence, the information required to hold the entropy increase to $\Delta H_{\rm tol}$—is given by

$$\Delta I_{\rm min} \sim \mathcal{R}(t)(\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}) \tag{6.6}$$

[cf. Eq. (5.35)], corresponding to there being a total of $(2^{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}})^{\mathcal{R}(t)}$ coarse-grained densities. The entropy increase (6.5) comes from counting the number of *non-overlapping* coarse-grained densities that are required to fill the volume occupied by the average density, that number being $2^{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}}$. In contrast, the information $\Delta I_{\rm min}$ comes from counting the exponentially greater number of ways of forming *overlapping* coarse-grained densities by choosing one of the $2^{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}}$ non-overlapping coarse-grained sub-densities in each of the $\mathcal{R}(t)$ correlation cells.

The picture developed in this section, summarized neatly in Eq. (6.6), requires that $\Delta H_{\rm tol}$ be big enough that a coarse-grained sub-density is much larger than a perturbed sub-density, so that we can talk meaningfully about the perturbed sub-densities that lie predominantly *within* a coarse-grained sub-density. If $\Delta H_{\rm tol}$ becomes too small, Eq. (6.6) breaks down, and the information $\Delta I_{\rm min}$, rather than reflecting a property of the chaotic dynamics as in Eq. (6.6), becomes essentially a property of the perturbation, reflecting a counting of the number of possible realizations of the perturbation.

The boundary between the two kinds of behavior of $\Delta I_{\rm min}$ is set roughly by the number $F$ of contracting phase-space dimensions. When $\Delta H_{\rm tol}/F \gtrsim 1$, the characteristic scale of a coarse-grained sub-density in the contracting dimensions is a factor of

$$(2^{\Delta H_{\rm tol}})^{1/F} = 2^{\Delta H_{\rm tol}/F} \gtrsim 2 \tag{6.7}$$

larger than the characteristic size of a perturbed sub-density in the contracting dimensions. In this regime the picture developed in this section is at least approximately valid, because a coarse-grained sub-density can accommodate several perturbed sub-densities in each contracting dimension. The information $\Delta I_{\rm min}$ becomes a property of the system dynamics, rather than a property of the perturbation, because it quantifies the effects of the perturbation on scales as big as or bigger than the finest scale set by the system dynamics.

In contrast, when $\Delta H_{\rm tol}/F \lesssim 1$, we are required to keep track of the phase-space density on a very fine scale in the contracting dimensions, a scale smaller than the characteristic size of a perturbed sub-density in the contracting dimensions. Sub-densities are considered to be distinct, even though they overlap substantially, provided that they differ by more than this very fine scale in the contracting dimensions. The information $\Delta I_{\rm min}$ is the logarithm of

the number of realizations of the perturbation which differ by more than this very fine scale in at least one correlation cell. The information becomes a property of the perturbation because it reports on the effects of the perturbation on scales finer than the finest scale set by the system dynamics—i.e., scales that are essentially irrelevant to the system dynamics.

We are now prepared to put in final form the exponential hypersensitivity to perturbation of systems with a positive KS entropy:

$$\frac{\Delta I_{\rm min}}{\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}} \sim \mathcal{R}(t) \sim 2^{K(t-t_0)} \quad \text{for } \Delta H_{\rm tol} \gtrsim F. \tag{6.8}$$

Once the chaotic dynamics renders the perturbation effective, this exponential hypersensitivity to perturbation is essentially independent of the form and strength of the perturbation. Its essence is that within each correlation cell there is a roughly even trade-off between entropy reduction and information, but for the entire phase-space density the trade-off is exponentially unfavorable because the density occupies an exponentially increasing number of correlation cells, in each of which it is perturbed independently.

As noted above, the behavior of $\Delta I_{\rm min}$ for $\Delta H_{\rm tol} \lesssim F$ deviates from the universal behavior of Eq. (6.8) and tells one about the number of realizations of the perturbation that produce densities that differ on scales finer than the finest scale set by the system dynamics. For a diffusive perturbation of the sort contemplated in this section, $\Delta I_{\rm min}$ diverges as $\Delta H_{\rm tol}$ goes to zero, because a diffusive perturbation has an infinite number of realizations on even the tiniest scale. If the diffusive perturbation is replaced by a similar perturbation, but with a finite number of realizations, then the growth of $\Delta I_{\rm min}$ is capped at the logarithm of the number of realizations, corresponding to the finest scale on which the perturbation acts. The perturbation used in the symbolic-dynamics analysis of perturbed chaotic maps in Sec. V is of this latter sort, with a finite number of realizations, the number being $D^{\mathcal{R}_n} = (2^{nh})^{\mathcal{R}_n}$. Indeed, the major simplifying assumption about the perturbation in Sec. V is that the sub-patterns produced by the perturbation are all different on the finest scale set by the system dynamics; i.e., there are no overlapping perturbed sub-patterns. This means that the cap on $\Delta I_{\rm min}$, which occurs at $\Delta I_{\rm min} \simeq \log_2(D^{\mathcal{R}_n}) = \mathcal{R}_n nh$ [cf. Eq. (5.16)], is such that the universal behavior of Eq. (5.35) extends right down to $\Delta H_{\rm tol} \simeq 0$.

What about systems with regular, or integrable dynamics? Though we expect no universal behavior for regular systems, we can get an idea of the possibilities from the heuristic description developed in this section. Hypersensitivity to perturbation requires, first, that the phase-space density develop structure on the scale of the strength of the perturbation, so that the perturbation becomes effective, and, second, that after the perturbation becomes effective, the phase-space density spread over many correlation cells.

For many regular systems there will be no hypersensitivity simply because the phase-space density does not develop fine enough structure. Regular dynamics can give rise to nonlinear shearing, however, in which case the density can develop structure on the scale of the strength of the perturbation and can spread over many correlation cells. In this situation, one expects the picture developed in this section to apply at least approximately: to hold the entropy increase to $\Delta H_{\rm tol}$ requires giving $\Delta H_{\mathcal{S}} - \Delta H_{\rm tol}$ bits per occupied correlation cell; $\Delta I_{\rm min}$ is related to $\Delta H_{\rm tol}$ by Eq. (6.6), with $\mathcal{R}(t)$ being the number of correlation cells occupied at time $t$. Thus regular systems can display hypersensitivity to perturbation if $\mathcal{R}(t)$ becomes large (although this behavior could be eliminated by choosing correlation cells that

are aligned with the nonlinear shearing produced by the system dynamics), but they cannot display *exponential* hypersensitivity to perturbation because the growth of $\mathcal{R}(t)$ is slower than exponential.

A more direct way of stating this conclusion is to reiterate what we have shown in this paper: Exponential hypersensitivity to perturbation is equivalent to the spreading of phase-space densities over an exponentially increasing number of phase-space cells; such exponential spreading holds for chaotic, but not for regular systems and is quantified by a positive value of the Kolmogorov-Sinai entropy.

## ACKNOWLEDGMENTS

The authors profited from discussions with H. Barnum and R. Menegus.

## REFERENCES

* Present Address: Department of Physics, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK. Email R.Schack@qmw.ac.uk.

[1] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion* (Springer, New York, 1983), pp. 268–275.
[2] V. M. Alekseev and M. V. Yakobson, Phys. Rep. **75**, 287 (1981).
[3] J. Ford, Phys. Today **36**(4), 40 (1983).
[4] B. O. Koopman, Proc. Nat. Acad. Sci. USA **17**, 315 (1931).
[5] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, The Netherlands, 1993).
[6] W. H. Zurek, Phys. Rev. A **40**, 4731 (1989).
[7] C. M. Caves, in *Complexity, Entropy, and the Physics of Information*, edited by W. H. Zurek (Addison-Wesley, Redwood City, CA, 1990), p. 91.
[8] C. M. Caves, in *Physical Origins of Time Asymmetry*, edited by J. J. Halliwell, J. Pérez-Mercader, and W. H. Zurek (Cambridge University Press, Cambridge, England, 1993), p. 47.
[9] C. M. Caves, Phys. Rev. E **47**, 4010 (1993).
[10] R. Schack and C. M. Caves, Phys. Rev. Lett. **69**, 3413 (1992).
[11] R. Schack and C. M. Caves, Phys. Rev. Lett. **71**, 525 (1993).
[12] R. Schack, G. M. D'Ariano, and C. M. Caves, Phys. Rev. E **50**, 972 (1994).
[13] M. V. Berry, in *New Trends in Nuclear Collective Dynamics*, edited by Y. Abe, H. Horiuchi, and K. Matsuyanagi (Springer, Berlin, 1992), p. 183.
[14] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948). Reprinted in book form, with postscript by W. Weaver: C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1949).
[15] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[16] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[17] E. T. Jaynes, in *Papers on Probability, Statistics, and Statistical Physics*, edited by R. D. Rosenkrantz (Kluwer, Dordrecht, Holland, 1983).
[18] H. Spohn, *Large Scale Dynamics of Interacting Particles* (Springer, Berlin, 1991).
[19] E. Borel, *Le Hasard* (Félix Alcan, Paris, 1914).
[20] E. Borel, *Traité du calcul des probabilités et de ses applications. Tome 2, Fasc. 3: Mécanique statistique classique* (Gauthier-Villars, Paris, 1925).
[21] R. Landauer, IBM J. Res. Develop. **5**, 183 (1961).
[22] R. Landauer, Nature **355**, 779 (1988).
[23] G. J. Chaitin, *Algorithmic Information Theory* (Cambridge University Press, Cambridge, England, 1987).
[24] W. H. Zurek, Nature **341**, 119 (1989).
[25] R. Schack, submitted to Phys. Rev. E (e-print hep-th/9409022).
[26] V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).
[27] D. S. Ornstein and B. Weiss, Am. Math. Soc. Bull. **24**, 11 (1991).
[28] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **119**, 861 (1958).
[29] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **124**, 754 (1959).
[30] J. P. Crutchfield and N. H. Packard, Physica **7D**, 201 (1983).

[31] D. Welsh, *Codes and Cryptography* (Clarendon Press, Oxford, 1988).